

*Citation for published version:*

Garzonis, S, Jones, S, Jay, T & O'Neill, E 2009, Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. in S Greenberg, SE Hudson, K Hinkley, M RingelMorris & DR Olsen (eds), CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, pp. 1513-1522, 27th Annual CHI Conference on Human Factors in Computing Systems, Boston, MA, USA United States, 4/04/09. <https://doi.org/10.1145/1518701.1518932>

DOI:

[10.1145/1518701.1518932](https://doi.org/10.1145/1518701.1518932)

Publication date:

2009

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Publisher Rights

Unspecified

© ACM, 2009. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Garzonis, S, Jones, S, Jay, T & O'Neill, E 2009, 'Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference'. in S Greenberg, SE Hudson, K Hinkley, M RingelMorris & DR Olsen (eds), CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery (ACM), pp. 1513-1522, 27th Annual CHI Conference on Human Factors in Computing Systems, Boston, MA, United States, 4-9 April <http://doi.acm.org/10.1145/1658550.1658558>

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference

Stavros Garzonis

Department of
Computer Science,
University of Bath,
Bath, BA2 7AY, UK
S.Garzonis@bath.ac.uk

Simon Jones

Department of
Computer Science,
University of Bath,
Bath, BA2 7AY, UK
S.Jones2@bath.ac.uk

Tim Jay

Graduate School of
Education,
University of Bristol
Bristol, BS8 1JA, UK
Tim.Jay@bristol.ac.uk

Eamonn O'Neill

Department of
Computer Science,
University of Bath,
Bath, BA2 7AY, UK
eamonn@cs.bath.ac.uk

ABSTRACT

With an ever increasing number of mobile services, meaningful audio notifications could effectively inform users of the incoming services while minimising undesired and intrusive interruptions. Therefore, careful design of mobile service notification is needed. In this paper we evaluate two types of audio (auditory icons and earcons) as mobile service notifications, by comparing them on 4 measures: intuitiveness, learnability, memorability and user preference. A 4-stage longitudinal evaluation involving two lab experiments, a field study and a web-based experiment indicated that auditory icons performed significantly better in all measures. Implications for mobile audio notification design are presented.

Author Keywords

Mobile audio notifications, mobile services, earcons, auditory icons, intuitiveness, learnability, memorability.

ACM Classification Keywords

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6)

INTRODUCTION

As mobile network providers are taking advantage of the available bandwidth and providing more data services, users are often left unaware of what services exist and when or where they are actually available. One obvious way to improve service awareness is notifying users of the availability of services relevant to their current situation or needs. For example, many users today have subscribed to receive alerts relating to sports results, weather forecasts, business news etc. Furthermore, as services become more context-aware, they can be tailored to the location of users (e.g. points of interest around them), their temporal dimension (e.g. events on specific dates and times) and their personal preferences.

However, audio notifications that appear too frequently can cause undesired interruptions, annoyance and ultimately their deactivation by the user [3]. Regardless of the service being suggested, users have to shift attention from the task at hand (e.g. conversing, driving or working) and read the message on their device, sometimes only to discover that they are not interested in the particular service suggested. An approach to address the trade-off between an unwelcome interruption and missing a potentially useful service is to improve the meaning of the audio notifications.

Designing meaningful audio notifications for mobile services is a challenge that has only recently been addressed [e.g. 11]. Meaning is attributed to sounds only through learning, whether explicit or implicit (perhaps with the exception of hard-wired semantics that have been established due to their importance for the survival of the species, e.g. the interpretation of the sound of a baby crying or of a loud bang). Therefore, if we wish to design meaningful notifications, we need to investigate their innate meaning (acquired either through evolutionary hard-wiring or cultural, implicit learning) and/or their ease of explicit learning and remembering. Furthermore, since interaction with mobile devices constitutes an everyday activity for most people and often takes place in social contexts, user preference in the audio types utilised is also an important factor to be considered in the notifications' design.

Within the realm of auditory display research, sounds have been classified in several types and their characteristics thoroughly described. Two common types are earcons and auditory icons. To decide which type of sound is more appropriate for supporting service awareness, we designed and carried out a 4-stage evaluation process, comparing certain cognitive attributes of the two types of sound. In particular, we investigated the immediacy of recognition of the notifications and their relation to the services (intuitiveness), the ease with which they can be learned to represent these services (learnability) and the ease with which these associations can be retained (memorability). Finally, user preferences were recorded throughout the evaluation and compared with the findings on the cognitive attributes of the sounds.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2009, April 4–9, 2009, Boston, MA, USA.

Copyright 2009 ACM 978-1-60558-246-7/08/04...\$5.00

In the remainder of this paper we present the evaluation process and discuss the findings. The next section presents relevant research on the types of sounds utilised and is followed by the presentation of the research that led to the mobile service categorisation used here. The following sections describe the design of the evaluation: a lab experiment for measuring intuitiveness and setting a baseline for measuring learnability; a field study for initiating and monitoring the learning process for the sound-service associations in a natural context; a second lab experiment for measuring learnability in the field (by comparing to the baseline from the first lab experiment) and learnability in the lab (through lab training); and a web-based experiment for measuring memorability of the associations after 1 and 4 weeks. Results, discussion and conclusions are presented towards the end of the paper.

BACKGROUND

Several types of sounds have been used in auditory interfaces. These may vary greatly in their semantic relationship with the events or situations they represent. This representational value is a continuum (e.g. from an arbitrary ‘beep’ sound to a meaningful speech message), and the directness of these mappings is believed to affect their ease of learning and retention [19]. Although speech is the most semantically rich acoustic medium, it also has a few shortcomings, one of which is privacy concerns, which can make it the least preferred option for some types of public notifications or reminders [e.g. 1, 22].

In this paper, we present an investigation into non-speech notifications for mobile devices, comparing the suitability of two widely acknowledged sound types in the literature: auditory icons and earcons.

Earcons

Blattner et al. [2] define earcons as “nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation or interaction”, while Brewster et al. [7] extended their description to “abstract, musical tones that can be used in structured combinations to create auditory messages”, “composed of short, rhythmic sequences of pitches with variable intensity, timbre and register”. Earcons are therefore flexible and can be designed to aurally extend any object, operation or interaction. Furthermore, they can be designed in families so that they represent hierarchies, by controlling or manipulating their different parameters (e.g. timbre and pitch). For example, Leplatre and Brewster [20] represented with earcons the hierarchical menu of a mobile phone, improving the performance of navigational tasks within the menu. On the flip side of flexibility, the major disadvantage of earcons is the lack of meaningful relationship with their referent. Users will usually have to learn and memorise these relationships from scratch, although some metaphorical value can sometimes be applied to display dimensions such as pitch [e.g. 24, 27].

Auditory Icons

‘Everyday listening’, the concept that “instead of mapping information to *sounds*, we can map information to *events*” was introduced by Gaver [12]. Unlike earcons, auditory icons utilise metaphors to relate them to their virtual referents, so that “if a good mapping between a source of sound and a source of data can be found, the meaning of an auditory icon should be easily learned and remembered” [14]. For example, the sound of shattering dishes can represent the drop of a virtual object into the (virtual) recycle bin [13], a door slamming indicates a remote user logging off the network [9], and tyre-skidding is used in vehicle collision warnings [15]. However, auditory icons lack flexibility, as metaphoric mappings are not always easy to find. A further disadvantage is that they can be confused with actual environmental sounds [9].

Both earcons and auditory icons can vary on the level of directness of the semantic relationship between the sounds and their referents. Nevertheless, it has been noted that users will construct their own meaning and stories to explain and memorise relationships between earcons [5] or auditory icons [9] and their referents.

Comparisons

Auditory notifications are usually researched in the context of safety-critical and/or cognitively demanding systems. Auditory icons and earcons have been utilised and their effectiveness as warning signals compared in power plants [16], operating theatres [24], collision avoidance systems in aircraft cockpits [23, 26] and in motor vehicles [15], while other studies have directly compared their learnability and memorability [e.g. 4, 10].

Auditory icon notifications are generally found to be easier to learn and retain [4, 10, 21, 26] and produce quicker reactions than earcon notifications [8, 15]. This superiority of auditory icons seems to be rooted in the directness of the association with their referents [19] and there is evidence suggesting that memory performance varies more on a sound-by-sound basis rather than a sound type-by-sound type basis [10]. Also, the type of training method has been found to affect learning and remembering audio notifications, regardless of the type of sound used [5, 10].

Finally, both earcons [18] and auditory icons [9] have been reported as annoying, but there are not many studies that measure and present comparative user preference findings. Two exceptions can be found in [25] where musical sounds were found more pleasant and appropriate than real world sounds, and [11] where findings regarding user preferences were inconclusive. It would appear that the context of use, the specific sound instances and individual aesthetic preferences outweigh any preference between the two sound types. We are not aware of previous longitudinal studies comparing user preference in the real context of use.

MOBILE SERVICE CATEGORISATION

Since the number of individual mobile services can be very large, it would be impossible to create meaningful and distinguishable non-speech audio notifications for each one of them. Therefore, in order to assess the suitability of the different audio types as mobile service notifications, a comprehensive categorisation of the services is needed. This way, each category can be assigned a unique notification, informing the user of the generic nature of the service rather than the specific instance of the category.

A previous attempt at mobile service categorisation was presented in [11] and was based on the concept of the services' origins. Three super-categories were devised, depending on whether the service was offered to the user from the 'world', from other users or from the user herself. Each category was further broken down into three subcategories and these nine categories were assigned individual audio notifications.

This categorisation has now been empirically validated and improved based on two card-sorting studies. The first study presented 10 groups of 3 or 4 participants with 52 services (an aggregation of the services of 3 major mobile networks in the UK), and participants were asked to group them so that each category would share the same notification. Cluster analysis was applied and the results indicated a categorisation very similar to the one devised in [11] with a few exceptions (e.g. some services were grouped together on the basis that they should only be accessible by adults). The clusters were slightly amended so that the hierarchy more accurately represented the origin-of-service rationale, and was then tested in a subsequent closed card-sorting study. This time the category titles and descriptions were provided to 12 participants, who were individually asked to assign each service to one of the categories. The categorisation was validated with participants reaching 83% agreement on average in assigning the services to the intended categories. Based on the qualitative results from the closed card-sorting study, some of the names and descriptions of the categories were amended, resulting in the hierarchy presented in Table 1. We used this categorisation in the study reported in this

A	Services from the world	Information Services	1	News Information
			2	Sports Information
			3	Here & Now Information
		Entertainment Services	4	Entertainment Downloads
			5	Entertainment Live
B	Services from other users	6 Incoming Calls		
		7 Incoming Messages		
C	Services from 'myself'	8 Self Reminders		
		9 Backup Reminders		
D	Other services	10 Other Services		

Table 1: Hierarchy of mobile services

paper to assign sound notifications to services. The following section describes our 4-stage evaluation study comparing the suitability of earcons and auditory icons as audio notifications for these mobile services.

EVALUATION STUDY

The evaluation study was carried out in 4 stages: a lab experiment, a field study, a second lab experiment and a web-based experiment. Next, we present the method of the study, including the procedures for each of the 4 stages.

Method

Design

A repeated measures experimental design was followed. The independent variables were time and training, and the dependent variables were response accuracy and response time. In the first stage (Lab 1), the intuitiveness of the notifications was measured. Intuition may be defined as "the act or faculty of knowing or sensing without the use of rational processes, or immediate cognition"¹. Therefore, the more immediately one recognises a sound and relates it to a concept or event, the more intuitive the sound's meaning. Since audio notifications' intuitiveness varies according to the strength of the metaphor associating them to their referents, auditory icons should (by definition) be more intuitive than earcons. Lab 1 had 3 goals: first, to test this assumption; secondly, to validate the strength of the metaphors of the auditory icons we used; thirdly, to get a baseline (i.e. pre-training) performance measure against which to compare subsequent (post-training) performance measures. In the second stage (Field Study), participants started learning the sound-service associations (while learnability was monitored) in the natural context of mobile phone usage, during a 1 week long field study. They received the audio notifications on their phones and they had to guess the corresponding service. They then received feedback messages indicating (or confirming if they got it right) the correct service. In related literature this learning process usually takes the form of training sessions in the lab, where participants are presented with stimuli as many times as needed for them to learn their meaning. This method might be appropriate for interfaces and systems where training is necessary (e.g. aircraft cockpits) but it is not how we normally

	Training	Day	Measurement	Hypotheses
Lab 1	No training	1	Intuitiveness	H1
Field Study	Field training	2-8	Learnability	H2
Lab 2	Lab training	8	Learnability	H3-H4
Web 1 Web 2	No training	15, 36	Memorability	H5

Table 2: The experimental design

¹ Source: www.thefreedictionary.com, 10/09/08

learn to use everyday technology such as a mobile phone. In the third stage (Lab 2), and since participants had not reached 100% accuracy during the field study training, they repeated the Lab 1 process so that learnability during the field study could be measured. Then, they underwent rigorous lab training and were again tested in how well they learned the associations. This provided us with further learnability data. Stage 4 (Web 1 and Web 2) tested memorability after 1 and 4 weeks with no further training or exposure to any of the sounds.

User preference data were collected throughout all 4 stages. In a previous study [11] it has been observed that questions such as “would you like to have this feature on your phone” give inconclusive results when the only exposure to the feature is during a short lab study. We suggest here that if users are exposed to the different sounds over a longer period, and in the natural context of everyday mobile device use, they will develop a clearer and more confident opinion of which sounds they would prefer to adopt, if any.

Hypotheses

Since only auditory icons are designed to have strong semantic relationships with their signified entities, it was hypothesised that *auditory icons will be significantly more intuitive notifications than earcons* (H1) as untrained listeners will have no basis on which to guess the association between earcons and services.

Previous research [e.g. 4, 26] suggests that auditory icons are more learnable than earcons during lab training sessions and there is no evidence to suggest that field training would be dissimilar. Furthermore, it has been found that users perform poorly in learning earcon associations in the field [18]. Therefore, we hypothesised that *auditory icons will be significantly more learnable notifications during the field study than earcons* (H2). Also, we predicted that *auditory icons would be quicker and easier to learn during the subsequent lab training* (H3).

During the lab training, participants were given an explanation of how the earcons represented the structure of the service hierarchy (Table 1). The associations between the sounds and the services were presented, and participants were given time to learn them. Following the findings of the literature with regard to training [e.g. 10], we predicted that *both sound types would perform significantly better after lab training* (H4). Finally, literature suggests that auditory icons are easier to retain than abstract sounds [21]. Therefore, in the web-based experiment, 1 and 4 weeks after the lab training, we predicted that *auditory icons’ forgetting rate will be slower than earcons’* (H5).

Participants

Sixteen participants took part in all 4 stages of the evaluation process, with the exception of 2 participants who could not complete the web-based experiment. There were 11 males and 5 females and their average age was 39.2 with a large range from 14 to 71 years old. All participants were

part of a cohort of a larger project and have been given mobile phones and free cellular airtime for 3 years in exchange for their participation in studies such as the one described here. No extra incentive was offered apart from 3 Amazon vouchers (of £50, £20 and £10) for the 3 most responsive participants during the field study. They were all familiar with using a computer, had been using a mobile phone for at least 3 years and were familiar with their current phone for at least 1 year.

Instruments and measures

The lab studies were designed and run using the MediaLab software, on IBM desktop PCs with a standard 15” display. Logitech headphones were attached to each computer and user input was recorded by MediaLab through the use of the keyboard and the mouse of each PC. For the field study, a Java application was installed on participants’ Nokia N95 phones, which communicated with a server initiating the notifications and recording the responses. For the last stage, a web-based questionnaire was developed in HTML and PHP, which was accessed via participants’ home or work computers.

Ten auditory icons and 10 earcons were used throughout the 4 stages of the evaluation process, each corresponding to one of the mobile services presented in the categorisation in Table 1. Next, we present the research carried out in order to generate and assign each set of stimuli to the services.

For auditory icons, everyday sound-producing events were initially assigned to the services as a result of a brainstorming session with four HCI researchers. All ideas of events that would effectively represent each of the services were aggregated and sound instances were collected from royalty free websites or recorded by the authors. The identifiability of these sounds was tested through an online survey and convenience sampling of respondents, with on average 81.5 responses for each sound. This led to the elimination of the sounds that were least identifiable. Only the 3 most identifiable sounds were retained for each service. (These were all identified by at least 64% of the respondents.) Next, a second online survey was carried out in order to assign to each service the most representative sound. The three most identified sounds were presented for each service and 112 participants were asked to indicate the one they found most representative. This ensured that the metaphors used to link the sounds with the services were empirically validated.

Earcons were initially designed to represent the hierarchy and be distinguishable, following guidelines from the literature [7]. Each one of the 4 super-categories (A-D in Table 1) was represented with a different instrument/timbre, which is reported as “the most efficient parameter that can be used to distinguish sounds” [6]. Within each category a variety in rhythm complexity (pitch, number of notes and duration) was utilised to make the earcons as distinct as possible. The earcons were experimentally tested with 12 participants against a set of control tones (varying only in

1	News Information	BBC News ident	Piano – monophonic going up
2	Sports Information	Stadium crowd	Piano – monophonic going down
3	"Here and Now" Information	Public announcement at an airport	Piano – monophonic jumps up & down
4	Entertainment - Downloads	20th Century Fox	Piano – polyphonic going down
5	Entertainment - Live	Audience applauding (e.g. in a theatre)	Piano – polyphonic going up
6	Incoming Calls	Old-fashioned phone ringing	Flute – monophonic short notes repeating
7	Incoming SMS Messages	Message transmitted in Morse code	Flute – monophonic long notes going down
8	Self Reminders	Windows Mobile reminder	Vibraphone – monophonic – going down
9	Backup Reminders	Truck/lorry reversing	Vibraphone – polyphonic – going up
10	Other Services	Wind chimes	Violin – varying pitch, chords and single notes

Table 3: The 10 services and the corresponding sound descriptions

pitch). They were found to be significantly better in their ability to represent the structure of the classification of the services, and more distinguishable than the control tones. However, the 5 services of the first super-category (represented by piano earcons) were found to be difficult to distinguish and remember. Therefore, they were improved by a professional musician in order to be more pleasant and more distinguishable. For example, the monophonic/ polyphonic dimension was manipulated to make sub-categories more distinct (Table 3). Twelve new participants tested the new earcons in a second experiment but no significant difference in learnability was found compared to the old earcons. Examples of earcons are presented in Table 4.

Table 3 shows the 10 service categories along with descriptions of the auditory icons and earcons. All sounds were sampled at 44.1 KHz at 128kbps bit rate and were normalised at -90 dB. However, during the field study their quality had to be reduced to just 8 KHz sampling due to the memory limitations of the mobile phones.

Procedures

Stage 1: Lab 1 - Intuitiveness

Up to 8 participants at a time performed the experiment at adjacent computers in a quiet lab and wearing headphones. Participants were presented with a list of the 10 services on their screens (Table 3) and the 4 super-categories rationale was explained (Table 1). Note that 2 extra 'dummy' ser-

Service 1	
Service 8	

Table 4: Examples of earcons

vices were added to the original classification that had no sounds corresponding to them. They were "Incoming MMS message" in the second super-category, and "Generic Alarm" in the third super-category. This was done to disrupt the one-to-one relationship between sounds and services, and hence hinder users from deducing associations based on a process of elimination.

Participants repeated the same procedure for earcons and for auditory icons with a short break in between. (The order of presentation was counterbalanced.) In each procedure, all 10 sounds were played in a random order twice, with all 10 sounds played once before the second round started. All 12 services (10 with sounds assigned and 2 additional dummy services) in their 4 super-categories were continuously presented on screen throughout the process, with a function key represented next to each one. Each time a sound was heard, participants were required to indicate (as "quickly and accurately" as they could) which service they thought it corresponded to by pressing a function key from F1 to F12. Response time and correct/incorrect responses were recorded. Each sound was repeated up to 3 times at 1.5 second intervals or until a selection was made. Each trial was followed by a short questionnaire capturing user preferences with regard to each individual sound. Finally, they were asked to answer a few comparative questions between the two sound types. All questions were presented on screen and responses were captured using MediaLab.

Stage 2: Field Study - Learnability

The field study commenced on the morning after the first lab study was completed. The application on the participant's mobile phone played the experimental sounds at random times from 10am to 9pm every day for 1 week. In order to draw participant's attention to the task, an SMS-type notification was played, immediately followed by an option screen allowing him either to begin the task and hear the sounds, or dismiss the task until the next random time interval had elapsed. Their task was to guess the correct correlation between a sound and a service (similar to the Lab 1 experiment). The numbered list of 12 services was displayed on the phone's screen and they were required to type the corresponding number in a text box. An on-screen button provided the option to replay the sound. They were given feedback on their selection, informing them whether it was incorrect or correct, and the correct service for the sound was indicated (or confirmed if guessed correctly).

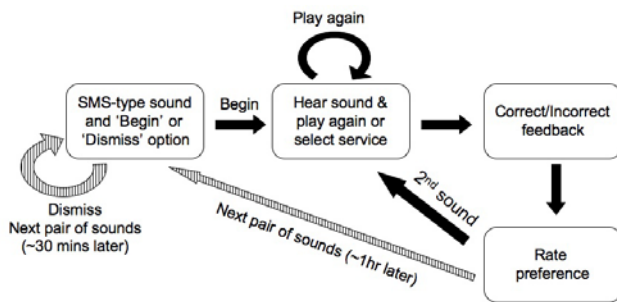


Figure 1: Mobile application interaction sequence

Finally, participants were asked to rate how much they would like to have that particular notification on their mobile phone for that service.

The sounds were presented in pseudo-random order in pairs: one earcon and one auditory icon (not necessarily for the same service) at a time. This was done for two reasons. First, it ensured that participants were responding to both types of sounds in the same context each time. As the participants' activities, level of concentration, distractions etc were expected to be changing throughout the day, we wanted to make sure that neither sound type was favoured by being responded to in more 'comfortable' situations. Secondly, the perceived annoyance or preference for the different sound types could also be affected by the context (especially the social context) within which the sounds were received. The presentation order of earcon and auditory icon within each pair was swapped each time.

Sounds that were responded to were not repeated in the same day. Pairs of sounds arrived approximately every 1 hour and each participant was expected to respond to all 20 sounds in every day (i.e. 10 pairs of sounds). If they failed to respond to some of the sounds, the frequency of the notifications was increased to around every half hour in order to increase the chances of responding to all the sounds within the day.

The interaction process of the mobile application is graphically modelled in Figure 1 and screenshots of the application are shown in Figure 2.

Stage 3: Lab 2 - Learnability

The Lab 2 experiment took place 1 week after Lab 1 (immediately after the field study). The first part of the Lab 2 experiment was identical to the evaluation process of the Lab 1 experiment. After participants made their (now in-

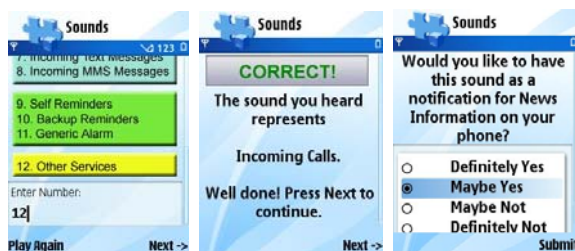


Figure 2: Mobile application screenshots

formed) guesses of corresponding service categories for both types of sounds, they were informed about the dummy services, which were then removed from the list.

A 4-step training process then took place for each of the sound types (order of presentation counterbalanced amongst participants). In the first step, each sound was presented along with its corresponding service, and their association was explained. For the auditory icons, the events that produced the sounds and the metaphors associating them to the services were described. For the earcons, the instrument and some properties of the musical pieces were described, along with any metaphorical associations (Table 3).

In the second step of training, participants were given 4 minutes to memorise the associations. The sound descriptions were displayed next to the services, and the sounds were played every time the participant clicked on the corresponding service buttons.

The third step of the training consisted of an absolute identification paradigm with trial-by-trial feedback. The 10 sounds were played in a random order and participants had to choose the correct service. They were prompted to retry if they made the wrong choice. Sounds were repeated up to 3 times or until the correct choice was made.

The final step of the training was a similar test but feedback was given only at the end of the trial of 10 sounds. Feedback consisted of a percentage score for their correct responses and a breakdown of which services they had guessed correctly. Training stopped when the participant scored 100% or had been through 6 repetitions of the whole process, whichever came first. The number of tests needed for each sound type was recorded to indicate ease of learnability for each sound type. The same training procedure was repeated for the second type of sound after a 5 minute break. A comparative questionnaire completed the experimental procedure. At the end of the experiment a short debriefing focus group took place, where participants were encouraged to describe their experiences during the lab and field sessions.

Stage 4: Web-based Experiments - Memorability

One and 4 weeks after the Lab 2 experiment, participants completed a web-based questionnaire. The procedure for the questionnaire was very similar to the fourth step of training in Lab 2. Participants listened to the 10 sounds of each type in a random order and indicated for each one the service they thought it represented. They had the chance to replay the sound and they were asked to indicate whether they "would like to receive this notification in the presence of others". At the end of each sound type they were provided with a percentage score for their correct responses, but no detailed breakdown.

RESULTS

The results for the study will be described in three parts. First, we present an analysis of participants' performance in

the two lab sessions. Secondly, we examine performance 1 and 4 weeks after the second lab session. Thirdly, we explore factors contributing to variance amongst participants' preference, and differences amongst notifications within the sound types.

Intuitiveness and Learnability

This section describes the analysis of participants' ability to correctly identify the meaning of notifications in the first lab session, plus the degree to which identification improved in the second lab session.

A 2x2 ANOVA was conducted, with two repeated measures variables, notification-type (auditory icon or earcon) and session (Lab 1 and Lab 2). There was a significant main effect of notification-type, $F=178.7$, $p<0.001$. This indicates that auditory icons were more easily associated with the correct service than were earcons regardless of training. A paired-samples t-test shows a significant difference between scores for the two notification types in the first session ($t(16)=11.007$, $p<0.001$). Thus the argument from definition that auditory icons are more intuitive than earcons is supported by the empirical evidence.

There was a significant main effect of session, $F=15.4$, $p=0.002$. This indicates that participants were more successful in correctly identifying services associated with both notification types in the Lab 2 session, after the field study training, than they were in Lab 1. There was a significant interaction between notification-type and session, $F=9.7$, $p=0.008$. This indicates that auditory icon associations were more easily learned during the week between lab sessions than were earcon associations. This is illustrated in Figure 3.

Reaction times showed similar characteristics to the number of correct identifications. Reaction times were lower for the identification of auditory icons, in both lab sessions, than for the identification of earcons, $F(14)=42.01$, $p<0.001$. Reaction times were also lower in the second lab session than the first, for both types of notification, $F(14)=49.05$, $p<0.001$. However, there was no significant interaction between the two, $F(14)=0.15$, $p=0.705$.

A t-test was conducted in order to compare the number of

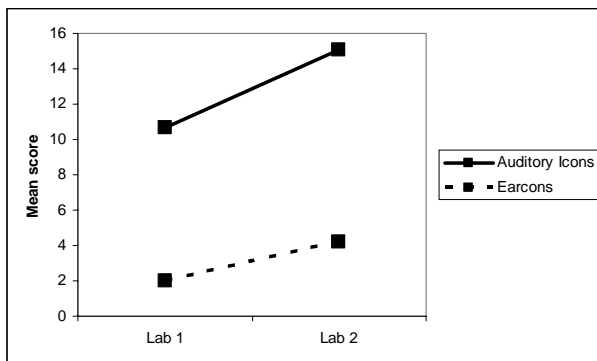


Figure 3: Change across lab sessions for both types of notification

trials required in the Lab 2 session to reach a 100% correct identification rate for each type of notification (see Table 6). There was a significant effect of notification type, $t=5.5$, $p<0.001$. This again indicates that it was easier for participants to learn associations between auditory icons and services than between earcons and services.

Forgetting and durability

Performance was analysed 1 and 4 weeks after the second lab session in order to determine whether one or other set of notifications was more quickly forgotten. A 3x2 ANOVA was conducted, with two repeated measures independent variables; notification-type (auditory icon and earcon) and session (Lab 2, Web 1 and Web 2). There was a significant main effect of time, $F(8)=14.13$, $p=0.002$. This indicates that associations between sound and meaning were forgotten over time, for both earcons and auditory icons, as would be expected. There was also a significant main effect of type, $F(9)=37.5$, $p<0.001$. This indicates that auditory icon associations were remembered better than were earcon associations at each of the 3 stages (Lab 2, Web 1 and Web 2). Finally, there was a significant interaction between time and type, $F(8)=5.07$, $p=0.038$. This shows that earcon associations were forgotten more quickly than were auditory icons. Figure 4 illustrates the overall change in performance of participants for each of the sessions described (note that the X-axis does not use a linear time scale). Sessions have been labelled according to the measurement they relate to (intuitiveness, learnability or memorability).

Preference and its effects amongst notifications

Looking at the four sessions together, there was a significant main effect of notification type on preference, $F(18)=19.72$, $p<0.001$, whereby participants preferred the auditory icons over the earcons. There was no significant

Notification type	Session	N	Mean	Std. Deviation
Auditory Icons	Lab 1	17	10.6471	2.84915
Earcons	Lab 1	17	2.0000	1.11803
Auditory Icons	Lab 2	15	15.0667	2.81493
Earcons	Lab 2	15	4.2000	3.12136

Table 5: Number of correct identifications (out of 20)

	N	Mean	Std. Deviation
Auditory Icons	15	1.1333	.51640
Earcons	15	3.7333	1.94447

Table 6: Number of trials required in lab 2 training to achieve 100% success

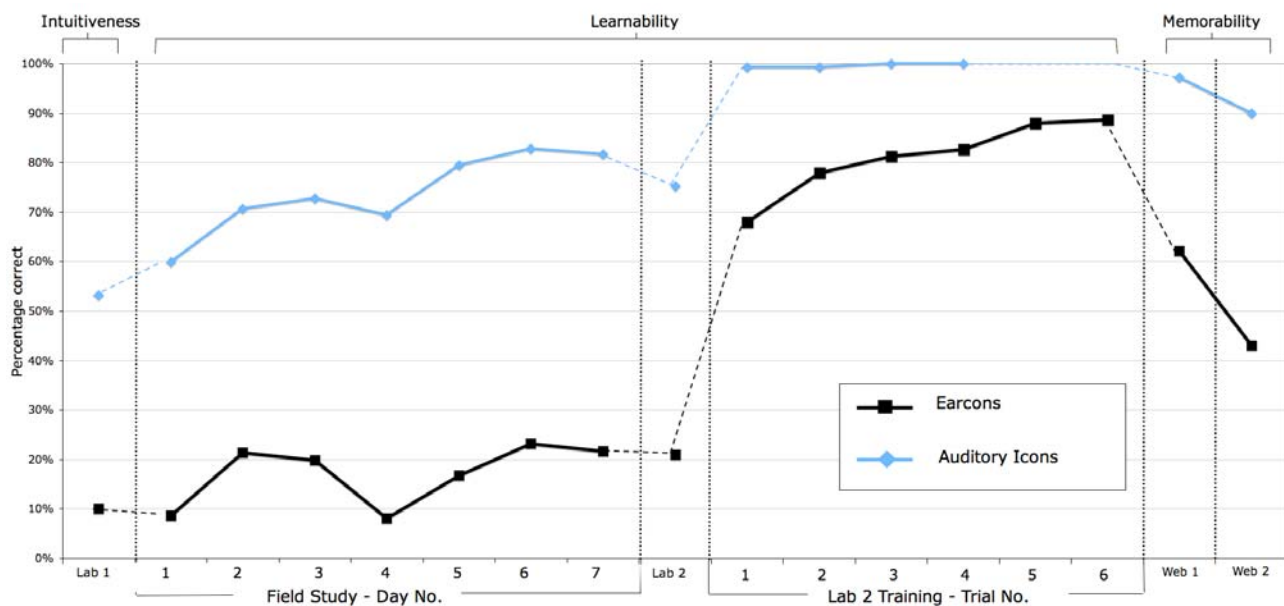


Figure 4: Percentage of correct identification in Lab 1, Lab 2, Web 1 and Web 2

main effect of session, $F(3)=1.96$, $p=0.13$, but there was a significant interaction effect (the difference in preference rating was largest in the Lab 2 session), $F(3)=11.00$, $p<0.001$.

Correlational analyses were used to compare average preference scores (from the first lab session) with performance in matching notifications with services in both the first and second lab sessions. There was a strong, positive correlation between preference and successful identification of notifications in both the first ($r=0.527$, $p=0.02$) and second lab sessions ($r=0.626$, $p=0.003$). This indicates that the most intuitive notifications were preferred more, and later on more successfully learned.

DISCUSSION

In all 4 stages of the study presented in this paper, several interesting results with regard to earcons' and auditory icons' suitability as mobile service notifications emerged. All of our hypotheses were supported with auditory icons performing significantly better in terms of intuitiveness (H1), learnability (H2, H3) and memorability (H5); and lab training significantly improved performance of both sound types (H4) but more for auditory icons (H3). Furthermore, auditory icons were consistently preferred over earcons as mobile audio notifications across all stages of the study.

Intuitiveness

Auditory icons performed significantly better than earcons in terms of intuitiveness, confirming the assumption from their respective definitions (H1), and validating our choice of metaphors for the auditory icons. However, the auditory icon for self-reminders (service 8) performed the worst in its group (9%), almost at chance rate. The sound chosen by the survey participants as the most representative for this

service was a standard Windows Mobile reminder notification. Such a sound might not be as widely known and recognised as the rest of the auditory icons and therefore could be perceived by some as an earcon (i.e. with arbitrary semantic connection to its service). This unfamiliarity is further underlined by the fact that participants in the Lab 1 experiment indicated at a rate of 56% that this sound actually represented an incoming SMS (which is higher than the 35% accuracy that the correct SMS sound achieved). The next 2 auditory icons to underperform were the 'truck reversing' (service 9) and the 'wind chimes' (service 10), scoring in Lab 1 at 21% each. Their performance increased to 100% right after the metaphors were explained during Lab 2 training. Therefore, we conclude that this poor performance was a result of poor choice of metaphors for these particular auditory icons. Indeed, in the identification survey the service 8 auditory icon achieved the lowest score of the accepted sounds, at only 64%. The next lowest identification score was for auditory icon 2 at 77%, which scored 97% in intuitiveness. This finding suggests that *sound-source identification rates have to be higher than 70% (across a convenient population sample) for successful mobile audio notification design.*

Learnability

An interesting result in terms of learnability is that earcons required only some basic training (presentation of the associations and 4 minutes exploration) in order to reach scores similar (but still lower) to the more intuitive auditory icons before training. Many participants reported with relief that earcons made much more sense after they were explained and the rationale for memorising them was presented. Some tried to create some semantic relationship between the earcons and the services while others based their learning purely on the order of presentation (first service had a

rising pitch, second had a falling pitch, third had a varying pitch etc). For participants to develop their own semantics or follow different strategies for memorising associations is not uncommon [4, 9]. However, the rate of improvement in their Lab 2 attempts to reach 100% accuracy was significantly slower than with auditory icons. Therefore, *basic training with earcons seems enough to achieve scores of about 70% but even after substantial training, they fail to reach auditory icons' performance*. It has been found that if more time is given for familiarisation earcon learnability improves [17], but for everyday technologies like the mobile phone, lengthy learning procedures are not ideal.

Memorability

Similar to the learnability results, earcons seem to be a less appropriate choice for mobile notifications, since their meaning is easier and quicker to forget. In a real world scenario the number of services could be more than just the 10 that were used here, but even more importantly, some of them could occur infrequently. Even if users undertook extensive learning to memorise the associations, if an earcon notification occurred once a week, the likelihood is that its meaning would have been forgotten. *Least frequent notifications have the greatest need to be based on meaningful metaphors*.

Preference

Apart from earcons scoring lower in subjective preference, they also seemed to cause strong negative feelings and frustration to some participants. This was apparent from involuntary comments during the training, explicit responses in the questionnaires and comments during the debriefing focus group after the Lab 2 experiment. For example, earcons were characterised by some as “horribly discordant”, “ugly” or “miserable” and one even stated “I loathe this sound” (earcon 10). However, a possible explanation for this negative attitude towards earcons is the inefficiency of the learning process during the field study. As one participant put it “Came to hate them all because I got them wrong so they just irritated me, which was a kind of vicious circle!” This is further supported by the strong correlation found between preference and intuitiveness. *Therefore, earcon design for mobile notifications needs not only to meet literature-based requirements in terms of structure and distinguishability, but also involve users in order to produce aesthetically pleasant sounds*. It is interesting to note that the same earcons received very positive feedback in the experiments establishing their validity (by comparing them with control sounds).

However, if one looks carefully at the preferences for individual sounds, there are earcons that scored relatively high. In particular, earcons 8 and 9 (vibraphone timbre) seem to score consistently higher than all other earcons in terms of preference throughout all stages of this study. This could possibly be due to the familiarity of the timbre, which sounded more similar to how mobile phones currently sound (in contrast for example to a piano). Also, there were

instances where participants found auditory icons to be “too intrusive” and potentially “embarrassing to broadcast loudly”. Because of these comments, we rephrased the preference question from “would you like to have this notification on your mobile phone” in the first 3 stages to “would you like to receive this notification in the presence of others” in the web-based experiments. However, the responses still favoured the auditory icons.

Finally, we have previously found [11] that preference questions with regard to potential real usage of audio types were inconclusive and unreliable when asked during a lab experiment. However, in the current study we found that participants were consistent in their preferences throughout the lab and field studies. We therefore may be able to put more trust in users' initial preference reactions to audio notifications.

Limitations

In the field study data, we observed a sudden drop in the performance of both types around day 4, and a slight drop on day 7 (Figure 4). This can be explained by a technical problem that significantly reduced the number of notifications delivered on day 4 and slightly reduced the responses on day 7. With fewer questions asked each answer had more effect on the average. As incorrect answers were more likely at that stage, the effect was to depress the averages. However, reduced responsiveness on Sunday also appeared for participants who did not face the technical problem, and similar but smaller unresponsiveness is also observed on Saturday and on a public holiday. Perhaps participants felt less willing to participate during their leisure time.

Furthermore, one of the limitations of the field application was that the sound quality had to be significantly reduced in order to run on the phones. It was observed (both by the researchers and some participants) that the acoustically more complex auditory icons were affected more than the earcons. Many participants reported that the low quality of the auditory icons made them annoying, in contrast to the high quality sounds during the lab sessions. Despite this quality bias in favour of earcons, participants still preferred the auditory icons during the field study.

CONCLUSION

In this paper we have presented an evaluation study comparing the appropriateness of auditory icons and earcons as mobile service notifications. Auditory icons performed significantly better in terms of intuitiveness, learnability, memorability and user preference. Drawing from these findings, we have discussed design implications for mobile audio notifications.

In particular, we suggest that commonly identified sounds should be used, or those that can be learned with quick and basic training. Also, less frequent notifications should be designed with stronger metaphors, and users should be involved in the notifications design process in order to avoid

negative feelings due to aesthetic preferences. Further research is needed to produce a more complete set of design guidelines for mobile audio notifications.

ACKNOWLEDGMENTS

This research was supported by the UK EPSRC, grant EP/C547683/1 "Cityware: urban design and pervasive systems". We thank musician James Mitchell for improving the earcon notifications, and Vodafone Group R&D and Nokia for their support of this research.

REFERENCES

1. Baber, C. and Noyes, J.M., Eds. (1993). *Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers*. Taylor & Francis, Inc.
2. Blattner, M.M., Sumikawa, D.A. and Greenberg, R.M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *SIGCHI Bull.* 21(1), pp. 123-124.
3. Block Jr F.E., Nuutinen L. and Ballast B., (1999). Optimization of Alarms: A Study on Alarm Limits, Alarm Sounds, and False Alarms, Intended to Reduce Annoyance. *Journal of Clinical Monitoring and Computing*, 15(2), pp. 75-83.
4. Bonebright, T.L., and Nees, M.A. (2007). Memory for Auditory Icons and Earcons with Localization Cues. In *Proc. ICAD 2007*, pp. 419-422.
5. Brewster, S.A. (1998). Using Non-speech Sounds to Provide Navigation Cues. *ACM Transactions on Computer-Human Interaction*, 5(2), pp. 224-259.
6. Brewster, S.A., Raty, V.P., and Kortekangas, A. (1996). Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy. In *Proc. BCS HCI'96*, pp. 169-183.
7. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. (1993). An Evaluation of Earcons for use in Auditory Human-Computer Interfaces. In *Proc. CHI'93*, pp. 222-227.
8. Bussemakers, M.P., and De Haan, A. (2000). When it Sounds Like a Duck and it Looks Like a Dog ... Auditory icons vs. earcons in multimedia environments. In *Proc. ICAD 2000*, pp. 184-189.
9. Cohen, J. (1994). Monitoring Background Activities. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification and Auditory interfaces*, pp. 499-522.
10. Edworthy, J. and Hards, R. (1999). Learning Auditory Warnings: The Effects of Sound Type, Verbal Labelling and Imagery on the Identification of Alarm Sounds. *International Journal of Industrial Ergonomics*, 24(6), pp. 603-618.
11. Garzonis, S., Bevan, C. and O'Neill, E. (2008). Mobile Service Audio Notifications: intuitive semantics and noises. In *Proc. OZCHI'08*, pp. 156-163.
12. Gaver, W.W. (1997). Auditory Interfaces. *Handbook of Human-Computer Interaction*, Elsevier Science.
13. Gaver, W.W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. In *Human-Computer Interaction*, 4(1), pp. 67-94.
14. Gaver, W.W. (1986). Auditory Icons: Using Sound in Computer Interfaces. In *Human-Computer Interaction*, 2(2), pp. 167-177.
15. Graham, R. (1999). Use of Auditory Icons as Emergency Warnings: Evaluation within a Vehicle Collision Avoidance Application. *Ergonomics*, 42(9), pp. 1233-1248.
16. Hickling, E. M. (1994). Ergonomics and Engineering Aspects of Designing an Alarm System for a Modern Nuclear Power Plant. In *Human Factors in Alarm Design*, N. Stanton, Ed. Taylor & Francis, pp. 165-178.
17. Hoggan, E. and Brewster, S.A. (2007). Designing Audio and Tactile Crossmodal Icons for Mobile Devices. In *Proc. ICMi'07*, ACM Press, pp. 162-169.
18. Isaacs, E., Walendowski, A. and Ranganathan, D. (2002). Hubbub: A Sound-Enhanced Mobile Instant Messenger that Supports Awareness and Opportunistic Interactions. In *Proc. CHI*, pp. 179-186.
19. Keller, P. and Stevens, C. (2004). Meaning from Environmental Sounds: Types of Signal-Referent Relations and their Effect on Recognizing Auditory Icons. *Journal of Experimental Psychology Applied*, 10(1), pp. 3-12.
20. Leplatre, G. and Brewster, S.A. (2000). Designing non-speech sounds to support navigation in mobile phone menus. In *Proc. ICAD'00*, pp. 190-199.
21. Leung, Y. K., Smith, S., Parker, S., and Martin, R. (1997). Learning and Retention of Auditory Warnings. In *Proc. ICAD'97*, pp. 288-299.
22. McGee-Lennon, M., Wolters M. and McBryan, T. (2007). Audio Reminders in the Home Environment. In *Proc ICAD'07*, pp. 437-444.
23. Patterson, R.D. and Mayfield, T.F. (1990). Auditory Warning Sounds in the Work Environment [and Discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 327(1241), *Human Factors in Hazardous Situations*, pp. 485-492.
24. Sanderson, P., Wee, A., Seah, E. and Lacherez, P., (2006). Auditory Alarms, Medical Standards, and Urgency. In *Proc. ICAD'06*, pp. 24-27.
25. Sikora, C.A., Roberts, L. and Murray, L. (1995). Musical vs. Real World Feedback Signals, In *Proc. CHI'95*, pp. 220-221.
26. Ulfvengren, P. (2003). Design of Natural Warning Sounds in Human-Machine Systems. PhD thesis, Stockholm, Sweden 2003.
27. Walker, B. N. and Kramer, G. (2005). Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Transactions on Applied Perception* 2(4), pp. 407-412.